

Machine Learning

Lab Assignment 3

(Linear Regression)

Note: This exercise is to solidify your understanding of basis function expansion, bias, variance, underfitting, overfitting, model capacity, regularization, ridge regression, and lasso regression.

Question 1

Let $f(x) = x + 2 \cdot \sin(0.5 \cdot x)$ be your true function.

1. **Generate the data:** Generate 50 data points for $x \in [-10, 10]$ by adding noise drawn from a Gaussian distribution with mean 0 and variance 0.5 to the true function.
2. **Optimal model capacity:** Use the learnt w to predict labels on test data and measure squared error on the test set. Plot the training and generalization error against the model capacity for each case.
3. **Addressing overfitting and Regularization:** Perform regularization on the curve with maximum overfitting.
 - **Use more data:** For the overfitting case, sample more training data points from the function, say 100, and now plot the true function, originally learnt curve, and the curve learnt on more data. What is the observation?
 - **Lasso Regression:** Perform Lasso regression on the cost function as follows, vary $\lambda = 0$ to 5 in steps of 0.5:

$$J(w) = \frac{1}{N} \sum_{i=1}^N (y_i - (w_0 + w^T \phi_n(x_i)))^2 + \lambda |w|$$

- Plot both training and test error for different values of λ or $(\ln \lambda)$ on the x-axis. What value of hyperparameter λ is most suitable? Is w sparse? Plot the number of non-zero components of w against λ .
 - Plot the true function, and curves learnt with and without regularization. Compare and comment.
 - Plot the L_1 norm of w against λ .
- **Ridge Regression:** Perform Ridge regression on the cost function as follows, vary regularization coefficient $\lambda = 0$ to 5 in steps of 0.5:

$$J(w) = \frac{1}{N} \sum_{i=1}^N (y_i - (w_0 + w^T \phi_n(x_i)))^2 + \lambda ||w||^2$$

- Plot both training and test error for different values of λ or $(\ln \lambda)$ on the x-axis. What value of hyperparameter λ is most suitable?
 - Plot the L_2 norm of w against λ .
 - Plot the true function, and curves learnt with and without regularization. Compare and comment.
- Does Lasso regression give sparser solution than Ridge regression (more number of zero components in w)?

Question 2

- a) Show the usage of [scikit-learn's linear regression module](#) for the [real estate price](#) prediction regression problem. What is the RMS error on the test set?
- b) Based on the regression coefficients, what can you comment about the importance of different features? Is it correct to assume that larger coefficients mean more important features?
- c) Now, standardize the dataset to have all features on a scale of 0 to 1. Re-learn the regression coefficients and now comment on the importance of different features.
- d) What is the distribution of the residuals?
- e) Use cross-correlation to find the optimal set of features to use for regression. Using all possible feature sets of length 1, 2, 3, or 4, what is the optimal feature set as per the validation set and how does this set of features perform on the test set wrt the model learned on the entire feature set?